

Exploring Octave package dataframe

April 10, 2019

Ge Baolai, *Western University*
SHARCNET | Compute Ontario

Topics

- Dataframe
- Installing package dataframe
- Accessing data in CSV files
- Manipulating data in a dataframe
- Other topics

General Interest Seminars 2019

```
octave:4> d1
d1 = dataframe with 32 rows and 7 columns
Src: mtcars.csv
```

_1	x__	mpg	cyl	disp	hp	drat	wt
Nr	char	double	double	double	double	double	double
1	Mazda RX4	21.000	6	160.000	110	3.9000	2.6200
2	Mazda RX4 Wag	21.000	6	160.000	110	3.9000	2.8750
3	Datsun 710	22.800	4	108.000	93	3.8500	2.3200
4	Hornet 4 Drive	21.400	6	258.000	110	3.0800	3.2150
5	Hornet Sportabout	18.700	8	360.000	175	3.1500	3.4400
6	Valiant	18.100	6	225.000	105	2.7600	3.4600
7	Duster 360	14.300	8	360.000	245	3.2100	3.5700
8	Merc 240D	24.400	4	146.700	62	3.6900	3.1900
9	Merc 230	22.800	4	140.800	95	3.9200	3.1500
10	Merc 280	19.200	6	167.600	123	3.9200	3.4400
11	Merc 280C	17.800	6	167.600	123	3.9200	3.4400
12	Merc 450SE	16.400	8	275.800	180	3.0700	4.0700
13	Merc 450SL	17.300	8	275.800	180	3.0700	3.7300
14	Merc 450SLC	15.200	8	275.800	180	3.0700	3.7800
15	Cadillac Fleetwood	10.400	8	472.000	205	2.9300	5.2500
16	Lincoln Continental	10.400	8	460.000	215	3.0000	5.4240
17	Chrysler Imperial	14.700	8	440.000	230	3.2300	5.3450
18	Fiat 128	32.400	4	78.700	66	4.0800	2.2000
19	Honda Civic	30.400	4	75.700	52	4.9300	1.6150
20	Toyota Corolla	33.900	4	71.100	65	4.2200	1.8350
21	Toyota Corona	21.500	4	120.100	97	3.7000	2.4650
22	Dodge Challenger	15.500	8	318.000	150	2.7600	3.5200
23	AMC Javelin	15.200	8	304.000	150	3.1500	3.4350
24	Camaro Z28	13.300	8	350.000	245	3.7300	3.8400
25	Pontiac Firebird	19.200	8	400.000	175	3.0800	3.8450
26	Fiat X1-9	27.300	4	79.000	66	4.0800	1.9350
27	Porsche 914-2	26.000	4	120.300	91	4.4300	2.1400
28	Lotus Europa	30.400	4	95.100	113	3.7700	1.5130
29	Ford Pantera L	15.800	8	351.000	264	4.2200	3.1700
30	Ferrari Dino	19.700	6	145.000	175	3.6200	2.7700
31	Maserati Bora	15.000	8	301.000	335	3.5400	3.5700
32	Volvo 142E	21.400	4	121.000	109	4.1100	2.7800

Dataframe



A data frame can be thought as an entity of the following

- A table of variables in columns and values in rows.
- Each column is a vector of data of certain type.
- There might be missing data, either empty or NA's.
- Associated with some common operations.



Example: Car data (from datasets in R)

_1	x__	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Nr	char	double	double	double	double	double	double	double	double	double	double	double
1	Mazda RX4	21.000	6	160.000	110	3.9000	2.6200	16.460	0	1	4	4
2	Mazda RX4 Wag	21.000	6	160.000	110	3.9000	2.8750	17.020	0	1	4	4
3	Datsun 710	22.800	4	108.000	93	3.8500	2.3200	18.610	1	1	4	1
4	Hornet 4 Drive	21.400	6	258.000	110	3.0800	3.2150	19.440	1	0	3	1
5	Hornet Sportabout	18.700	8	360.000	175	3.1500	3.4400	17.020	0	0	3	2
6	Valiant	18.100	6	225.000	105	2.7600	3.4600	20.220	1	0	3	1
7	Duster 360	14.300	8	360.000	245	3.2100	3.5700	15.840	0	0	3	4
8	Merc 240D	24.400	4	146.700	62	3.6900	3.1900	20.000	1	0	4	2
9	Merc 230	22.800	4	140.800	95	3.9200	3.1500	22.900	1	0	4	2
10	Merc 280	19.200	6	167.600	123	3.9200	3.4400	18.300	1	0	4	4
11	Merc 280C	17.800	6	167.600	123	3.9200	3.4400	18.900	1	0	4	4
12	Merc 450SE	16.400	8	275.800	180	3.0700	4.0700	17.400	0	0	3	3
13	Merc 450SL	17.300	8	275.800	180	3.0700	3.7300	17.600	0	0	3	3
14	Merc 450SLC	15.200	8	275.800	180	3.0700	3.7800	18.000	0	0	3	3
15	Cadillac Fleetwood	10.400	8	472.000	205	2.9300	5.2500	17.980	0	0	3	4
16	Lincoln Continental	10.400	8	460.000	215	3.0000	5.4240	17.820	0	0	3	4
17	Chrysler Imperial	14.700	8	440.000	230	3.2300	5.3450	17.420	0	0	3	4
18	Fiat 128	32.400	4	78.700	66	4.0800	2.2000	19.470	1	1	4	1
19	Honda Civic	30.400	4	75.700	52	4.9300	1.6150	18.520	1	1	4	2
20	Toyota Corolla	33.900	4	71.100	65	4.2200	1.8350	19.900	1	1	4	1
21	Toyota Corona	21.500	4	120.100	97	3.7000	2.4650	20.010	1	0	3	1
22	Dodge Challenger	15.500	8	318.000	150	2.7600	3.5200	16.870	0	0	3	2
23	AMC Javelin	15.200	8	304.000	150	3.1500	3.4350	17.300	0	0	3	2
24	Camaro Z28	13.300	8	350.000	245	3.7300	3.8400	15.410	0	0	3	4
25	Pontiac Firebird	19.200	8	400.000	175	3.0800	3.8450	17.050	0	0	3	2
26	Fiat X1-9	27.300	4	79.000	66	4.0800	1.9350	18.900	1	1	4	1
27	Porsche 914-2	26.000	4	120.300	91	4.4300	2.1400	16.700	0	1	5	2
28	Lotus Europa	30.400	4	95.100	113	3.7700	1.5130	16.900	1	1	5	2
29	Ford Pantera L	15.800	8	351.000	264	4.2200	3.1700	14.500	0	1	5	4
30	Ferrari Dino	19.700	6	145.000	175	3.6200	2.7700	15.500	0	1	5	6
31	Maserati Bora	15.000	8	301.000	335	3.5400	3.5700	14.600	0	1	5	8
32	Volvo 142E	21.400	4	121.000	109	4.1100	2.7800	18.600	1	1	4	2



Example (cont'd): Extracting certain columns

_1	x__	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Nr	char	double	double	double	double	double	double	double	double	double	double	double
1	Mazda RX4	21.000	6	160.000	110	3.9000	2.6200	16.460	0	1	4	4
2	Mazda RX4 Wag	21.000	6	160.000	110	3.9000	2.8750	17.020	0	1	4	4
3	Datsun 710	22.800	4	108.000	93	3.8500	2.3200	18.610	1	1	4	1
4	Hornet 4 Drive	21.400	6	258.000	110	3.0800	3.2150	19.440	1	0	3	1
5	Hornet Sportabout	18.700	8	360.000	175	3.1500	3.4400	17.020	0	0	3	2
6	Valiant	18.100	6	225.000	105	2.7600	3.4600	20.220	1	0	3	1
7	Duster 360	14.300	8	360.000	245	3.2100	3.5700	15.840	0	0	3	4
8	Merc 240D	24.400	4	146.700	62	3.6900	3.1900	20.000	1	0	4	2
9	Merc 230	22.800	4	140.800	95	3.9200	3.1500	22.900	1	0	4	2
10	Merc 280	19.200	6	167.600	123	3.9200	3.4400	18.300	1	0	4	4
11	Merc 280C	17.800	6	167.600	123	3.9200	3.4400	18.900	1	0	4	4
12	Merc 450SE	16.400	8	275.800	180	3.0700	4.0700	17.400	0	0	3	3
13	Merc 450SL	17.300	8	275.800	180	3.0700	3.7300	17.600	0	0	3	3
14	Merc 450SLC	15.200	8	275.800	180	3.0700	3.7800	18.000	0	0	3	3
15	Cadillac Fleetwood	10.400	8	472.000	205	2.9300	5.2500	17.980	0	0	3	4
16	Lincoln Continental	10.400	8	460.000	215	3.0000	5.4240	17.820	0	0	3	4
17	Chrysler Imperial	14.700	8	440.000	230	3.2300	5.3450	17.420	0	0	3	4
18	Fiat 128	32.400	4	78.700	66	4.0800	2.2000	19.470	1	1	4	1
19	Honda Civic	30.400	4	75.700	52	4.9300	1.6150	18.520	1	1	4	2
20	Toyota Corolla	33.900	4	71.100	65	4.2200	1.8350	19.900	1	1	4	1
21	Toyota Corona	21.500	4	120.100	97	3.7000	2.4650	20.010	1	0	3	1
22	Dodge Challenger	15.500	8	318.000	150	2.7600	3.5200	16.870	0	0	3	2
23	AMC Javelin	15.200	8	304.000	150	3.1500	3.4350	17.300	0	0	3	2
24	Camaro Z28	13.300	8	350.000	245	3.7300	3.8400	15.410	0	0	3	4
25	Pontiac Firebird	19.200	8	400.000	175	3.0800	3.8450	17.050	0	0	3	2
26	Fiat X1-9	27.300	4	79.000	66	4.0800	1.9350	18.900	1	1	4	1
27	Porsche 914-2	26.000	4	120.300	91	4.4300	2.1400	16.700	0	1	5	2
28	Lotus Europa	30.400	4	95.100	113	3.7700	1.5130	16.900	1	1	5	2
29	Ford Pantera L	15.800	8	351.000	264	4.2200	3.1700	14.500	0	1	5	4
30	Ferrari Dino	19.700	6	145.000	175	3.6200	2.7700	15.500	0	1	5	6
31	Maserati Bora	15.000	8	301.000	335	3.5400	3.5700	14.600	0	1	5	8
32	Volvo 142E	21.400	4	121.000	109	4.1100	2.7800	18.600	1	1	4	2



Example (cont'd): Extracting certain rows

_1	x__	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Nr	char	double	double	double	double	double	double	double	double	double	double	double
1	Mazda RX4	21.000	6	160.000	110	3.9000	2.6200	16.460	0	1	4	4
2	Mazda RX4 Wag	21.000	6	160.000	110	3.9000	2.8750	17.020	0	1	4	4
3	Datsun 710	22.800	4	108.000	93	3.8500	2.3200	18.610	1	1	4	1
4	Hornet 4 Drive	21.400	6	258.000	110	3.0800	3.2150	19.440	1	0	3	1
5	Hornet Sportabout	18.700	8	360.000	175	3.1500	3.4400	17.020	0	0	3	2
6	Valiant	18.100	6	225.000	105	2.7600	3.4600	20.220	1	0	3	1
7	Duster 360	14.300	8	360.000	245	3.2100	3.5700	15.840	0	0	3	4
8	Merc 240D	24.400	4	146.700	62	3.6900	3.1900	20.000	1	0	4	2
9	Merc 230	22.800	4	140.800	95	3.9200	3.1500	22.900	1	0	4	2
10	Merc 280	19.200	6	167.600	123	3.9200	3.4400	18.300	1	0	4	4
11	Merc 280C	17.800	6	167.600	123	3.9200	3.4400	18.900	1	0	4	4
12	Merc 450SE	16.400	8	275.800	180	3.0700	4.0700	17.400	0	0	3	3
13	Merc 450SL	17.300	8	275.800	180	3.0700	3.7300	17.600	0	0	3	3
14	Merc 450SLC	15.200	8	275.800	180	3.0700	3.7800	18.000	0	0	3	3
15	Cadillac Fleetwood	10.400	8	472.000	205	2.9300	5.2500	17.980	0	0	3	4
16	Lincoln Continental	10.400	8	460.000	215	3.0000	5.4240	17.820	0	0	3	4
17	Chrysler Imperial	14.700	8	440.000	230	3.2300	5.3450	17.420	0	0	3	4
18	Fiat 128	32.400	4	78.700	66	4.0800	2.2000	19.470	1	1	4	1
19	Honda Civic	30.400	4	75.700	52	4.9300	1.6150	18.520	1	1	4	2
20	Toyota Corolla	33.900	4	71.100	65	4.2200	1.8350	19.900	1	1	4	1
21	Toyota Corona	21.500	4	120.100	97	3.7000	2.4650	20.010	1	0	3	1
22	Dodge Challenger	15.500	8	318.000	150	2.7600	3.5200	16.870	0	0	3	2
23	AMC Javelin	15.200	8	304.000	150	3.1500	3.4350	17.300	0	0	3	2
24	Camaro Z28	13.300	8	350.000	245	3.7300	3.8400	15.410	0	0	3	4
25	Pontiac Firebird	19.200	8	400.000	175	3.0800	3.8450	17.050	0	0	3	2
26	Fiat X1-9	27.300	4	79.000	66	4.0800	1.9350	18.900	1	1	4	1
27	Porsche 914-2	26.000	4	120.300	91	4.4300	2.1400	16.700	0	1	5	2
28	Lotus Europa	30.400	4	95.100	113	3.7700	1.5130	16.900	1	1	5	2
29	Ford Pantera L	15.800	8	351.000	264	4.2200	3.1700	14.500	0	1	5	4
30	Ferrari Dino	19.700	6	145.000	175	3.6200	2.7700	15.500	0	1	5	6
31	Maserati Bora	15.000	8	301.000	335	3.5400	3.5700	14.600	0	1	5	8
32	Volvo 142E	21.400	4	121.000	109	4.1100	2.7800	18.600	1	1	4	2



Example (cont'd): Extracting a grid of cells

_1	x__	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Nr	char	double	double	double	double	double	double	double	double	double	double	double
1	Mazda RX4	21.000	6	160.000	110	3.9000	2.6200	16.460	0	1	4	4
2	Mazda RX4 Wag	21.000	6	160.000	110	3.9000	2.8750	17.020	0	1	4	4
3	Datsun 710	22.800	4	108.000	93	3.8500	2.3200	18.610	1	1	4	1
4	Hornet 4 Drive	21.400	6	258.000	110	3.0800	3.2150	19.440	1	0	3	1
5	Hornet Sportabout	18.700	8	360.000	175	3.1500	3.4400	17.020	0	0	3	2
6	Valiant	18.100	6	225.000	105	2.7600	3.4600	20.220	1	0	3	1
7	Duster 360	14.300	8	360.000	245	3.2100	3.5700	15.840	0	0	3	4
8	Merc 240D	24.400	4	146.700	62	3.6900	3.1900	20.000	1	0	4	2
9	Merc 230	22.800	4	140.800	95	3.9200	3.1500	22.900	1	0	4	2
10	Merc 280	19.200	6	167.600	123	3.9200	3.4400	18.300	1	0	4	4
11	Merc 280C	17.800	6	167.600	123	3.9200	3.4400	18.900	1	0	4	4
12	Merc 450SE	16.400	8	275.800	180	3.0700	4.0700	17.400	0	0	3	3
13	Merc 450SL	17.300	8	275.800	180	3.0700	3.7300	17.600	0	0	3	3
14	Merc 450SLC	15.200	8	275.800	180	3.0700	3.7800	18.000	0	0	3	3
15	Cadillac Fleetwood	10.400	8	472.000	205	2.9300	5.2500	17.980	0	0	3	4
16	Lincoln Continental	10.400	8	460.000	215	3.0000	5.4240	17.820	0	0	3	4
17	Chrysler Imperial	14.700	8	440.000	230	3.2300	5.3450	17.420	0	0	3	4
18	Fiat 128	32.400	4	78.700	66	4.0800	2.2000	19.470	1	1	4	1
19	Honda Civic	30.400	4	75.700	52	4.9300	1.6150	18.520	1	1	4	2
20	Toyota Corolla	33.900	4	71.100	65	4.2200	1.8350	19.900	1	1	4	1
21	Toyota Corona	21.500	4	120.100	97	3.7000	2.4650	20.010	1	0	3	1
22	Dodge Challenger	15.500	8	318.000	150	2.7600	3.5200	16.870	0	0	3	2
23	AMC Javelin	15.200	8	304.000	150	3.1500	3.4350	17.300	0	0	3	2
24	Camaro Z28	13.300	8	350.000	245	3.7300	3.8400	15.410	0	0	3	4
25	Pontiac Firebird	19.200	8	400.000	175	3.0800	3.8450	17.050	0	0	3	2
26	Fiat X1-9	27.300	4	79.000	66	4.0800	1.9350	18.900	1	1	4	1
27	Porsche 914-2	26.000	4	120.300	91	4.4300	2.1400	16.700	0	1	5	2
28	Lotus Europa	30.400	4	95.100	113	3.7700	1.5130	16.900	1	1	5	2
29	Ford Pantera L	15.800	8	351.000	264	4.2200	3.1700	14.500	0	1	5	4
30	Ferrari Dino	19.700	6	145.000	175	3.6200	2.7700	15.500	0	1	5	6
31	Maserati Bora	15.000	8	301.000	335	3.5400	3.5700	14.600	0	1	5	8
32	Volvo 142E	21.400	4	121.000	109	4.1100	2.7800	18.600	1	1	4	2



Some typical operations

- Extract a subset of data according to certain criteria, e.g. find all records of cars with 6 cylinders, or fuel economy $< 6\text{L}/100\text{km}$ hwy, etc.
- Find the mean, max/min of a column.
- Find the sum, mean, std, etc. of a column by group in another column.



Dataframe provides a convenient and efficient way for data access and manipulations

- Access columns by name, so the relative position of a column in a dataset is irrelevant.
- Operations are (normally) vectorized.
- Well supported in R, Python for long time.
- Supported in Matlab (data table) and Octave (through package **dataframe**).



Dataframe in Octave

- Supported through package dataframe.
- It is claimed to be an implementation of dataframe in R.
- It does most of the dataframe if R does. If one primarily works with Matlab/Octave and wants to use some of the common dataframe features, then the dataframe package for Octave might be good enough.
- The current version is 1.2, the documentation is minimum. One needs to have a good grasp of Octave language features, especially the vector operations and cell arrays.



Demo



Exercises:

- Installing dataframe package in Octave
- Loading CSV files into dataframe
- Creating a dataframe
- Accessing and manipulating data in a dataframe
- Aggregation
- Discussion

